

Examining Distributions

Checkpoint 2

Pool2

Question (1)

The distribution of the amount of money spent by students for textbooks in a semester is approximately normal in shape with a mean of \$235 and a standard deviation of \$20. According to the standard deviation rule, almost all (99.7%) of the students spent on textbooks in a semester:

- A:** between 215 and 255 dollars.
- B:** between 195 and 275 dollars.
- C:** between 175 and 295 dollars.
- D:** less than 215 dollars or more than 255 dollars.
- E:** above 235 dollars.

Feedback

A : 0

- X** This is not quite right. The Standard deviation rule tells us that for distributions that have the normal shape, approximately 99.7% of the observations fall within *three* standard deviations of the mean. Note that 215 and 255 are only *one* standard deviation below and above the mean, respectively. (C) is the correct answer.

B : 0

- X** This is not quite right. The Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 99.7% of the observations fall within *three* standard deviations of the mean. Note that 195 and 275 are only *two* standard deviations below and above the mean, respectively. (C) is the correct answer.

C : 10

✓ Good Job! The Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 99.7% of the observations fall within *three* standard deviations of the mean. Indeed, $175 = 235 - 3 \cdot 20$ and $295 = 235 + 3 \cdot 20$ are exactly three standard deviations below and above the mean, respectively.

D : 0

✗ This is not quite right. The Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 99.7% of the observations fall *within three* standard deviations of the mean. Note that 215 and 255 are only *one* standard deviation below and above the mean, respectively. (C) is the correct answer.

E : 0

✗ This is not quite right. Since 235 is the mean, roughly 50% of the observations fall above it. Recall that the Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 99.7% of the observations fall *within three* standard deviations of the mean. (C) is the correct answer.


Question (2)

The distribution of the amount of money spent by students for textbooks in a semester is approximately normal in shape with a mean of \$235 and a standard deviation of \$20. According to the standard deviation rule, almost 2.5% of the students spent more than _____ on textbooks in a semester.


A: \$195**B:** \$215**C:** \$235**D:** \$275**E:** \$295

Feedback


A : 0

-  That is not quite right. The Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 95% of the observations will fall within 2 standard deviations of the mean. In this case the mean is \$235, and the standard deviation is \$20; therefore, approximately 95% of the observations (amount of money spent by students on textbooks) fall between $\$235 - 2 * \20 and $\$235 + 2 * \20 or between \$195 and \$275. Approximately 5% of the observations fall outside this interval. Since the "normal" shape is symmetric, approximately 2.5% of the observations fall outside this on each end. Which extreme 2.5% were you asked to find—the upper or lower? (D) is the correct answer.


B : 0

-  That is not quite right. The Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 95% of the observations will fall within 2 standard deviations of the mean. In this case the mean is \$235, and the standard deviation is \$20; check your calculations. Did you include two standard deviations? Also, which extreme were you asked to find—the upper or lower? (D) is the correct answer.

C : 0

-  That is not quite right. The Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 95% of the observations will fall within 2 standard deviations of the mean. In this case, the mean is \$235, and the standard deviation is \$20. Can the mean be the end of this interval? (D) is the correct answer.

D : 10

-  Good job! The Standard Deviation Rule tells us that for distributions that have the normal shape, approximately 95% of the observations will fall within 2 standard deviations of the mean. In this case the mean is \$235, and the standard deviation is \$20; therefore, approximately 95% of the observations (amount of money spent by students on textbooks) falls between $\$235 - 2 * \20 and $\$235 + 2 * \20 or between \$195 and \$275. Approximately 5% of the observations fall outside this interval. Since the "normal" shape is symmetric, approximately 2.5% of the observations fall above \$275.

E : 0

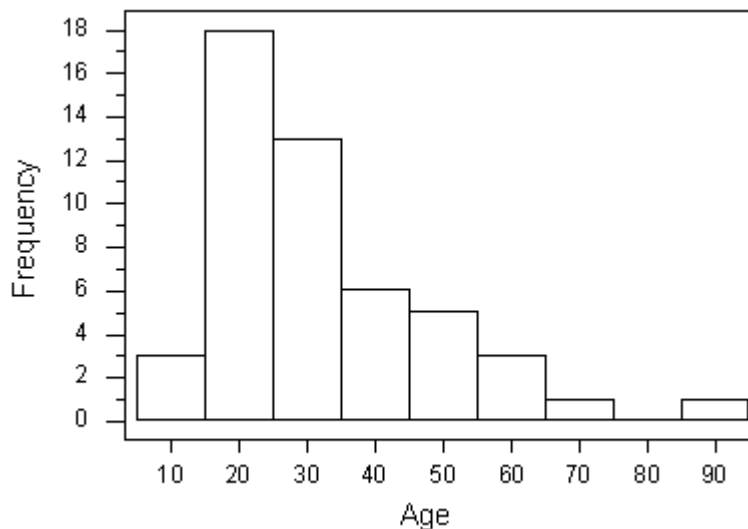
-  That is not quite right. The Standard Deviation Rule tells us that

X for distributions that have the normal shape, approximately 95% of the observations will fall within 2 standard deviations of the mean. In this case the mean is \$235, and the standard deviation is \$20; check your calculations. Did you include only two standard deviations? (D) is the correct answer.

Pool3

Question (3)

The histogram below displays the distribution of 50 ages at death due to trauma (unnatural accidents and homicides) that were observed in a certain hospital during a week.



To describe the center and spread of the above distribution, the appropriate numerical measures are:


- A:** the mean and the median
- B:** the mean and the standard deviation
- C:** the IQR and the standard deviation

D: the median and the IQR


E: Any of the above would be appropriate, it is just a matter of taste.

Feedback


A : 0

 This is not quite right. Remember that the mean and the median are both measures of **center**. But in this question, we want one measure of **center** and one measure of **spread**. (D) is the correct answer.


B : 0

 This is not quite right. Remember that the mean is only the best measure of center (and the standard deviation is only the corresponding best measure of spread) when the distribution is reasonably symmetric and has no outliers. But this distribution is skewed (it has a "tail" towards one side) and it also has a suspected outlier (the value at 90). (D) is the correct answer.


C : 0

 This is not quite right. Remember that the IQR and the standard deviation are both measures of **spread**. But in this question, we want one measure of **center** and one measure of **spread**. (D) is the correct answer.

D : 10

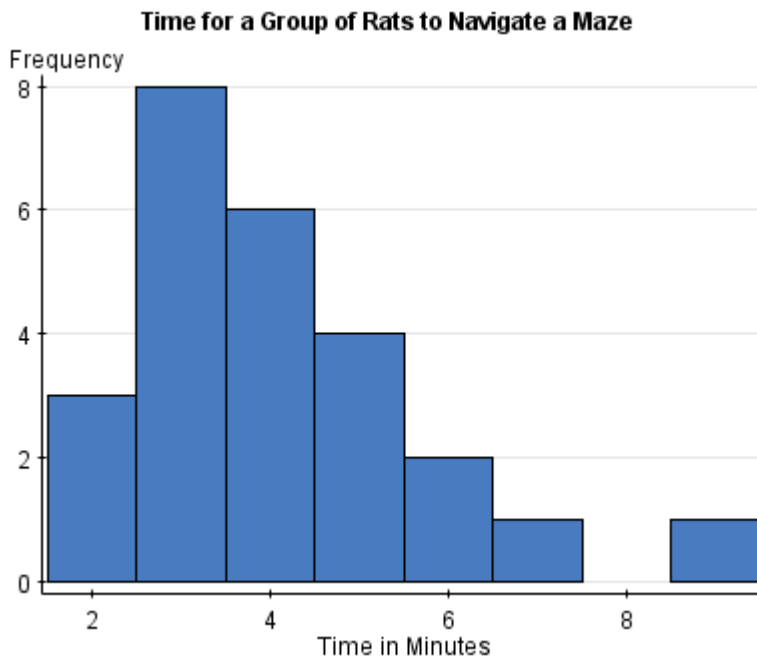
 Good job! When the distribution is skewed or has outliers, the best measure of center is the median, and the associated best measure of spread is the IQR. In this exercise, the distribution is skewed (it has a "tail" towards one side) and it also has a suspected outlier (the value at 90).

E : 0

 This is not quite right. Notice the distribution is skewed (it has a "tail" towards one side) and it also has a suspected outlier (the value at 90). Remember that when a distribution is skewed or has outliers, the median is the most appropriate measure of center, and the IQR is the most appropriate measure of spread. Using a measure of center other than the median, or using a measure of spread other than the IQR, would give misleading results for this distribution. (D) is the correct answer.

Question (4)

The histogram below shows the times, in minutes, required for 25 rats in a animal behavior experiment to successfully navigate a maze.



To describe the center and spread of the above distribution, the appropriate numerical measures are:


- A:** the mean and the median
- B:** the mean and the standard deviation
- C:** the IQR and the standard deviation
- D:** the median and the IQR
- E:** Any of the above would be appropriate, it is just a matter of taste.

Feedback


A : 0

X This is not quite right. Remember that the mean and the median are both measures of **center**. But in this question, we want one measure of **center** and one measure of **spread**. (D) is the correct answer.


B : 0

-  This is not quite right. Remember that the mean is only the best measure of center (and the standard deviation is only the corresponding best measure of spread) when the distribution is reasonably symmetric and has no outliers. But this distribution is skewed (it has a "tail" towards one side) and it also has a suspected outlier (the value at 9). (D) is the correct answer.


C : 0

-  This is not quite right. Remember that the IQR and the standard deviation are both measures of **spread**. But in this question, we want one measure of **center** and one measure of **spread**. (D) is the correct answer.

D : 10

-  Good job! When the distribution is skewed or has outliers, the best measure of center is the median, and the associated best measure of spread is the IQR. In this exercise, the distribution is skewed (it has a "tail" towards one side) and it also has a suspected outlier (the value at 9).

E : 0

-  This is not quite right. Notice the distribution is skewed (it has a "tail" towards one side) and it also has a suspected outlier (the value at 9). Remember that when a distribution is skewed or has outliers, the median is the most appropriate measure of center, and the IQR is the most appropriate measure of spread. Using a measure of center other than the median, or using a measure of spread other than the IQR, would give misleading results for this distribution. (D) is the correct answer.

Pool4

Question (5)

A student survey was conducted in a major university, where data were collected from a random sample of 750 undergraduate students. One variable that was recorded for each student was the student's answer to the question: "With whom do you find it easiest to make friends? Opposite sex/same sex/no difference."

These data would be best displayed using which of the following?

- A:** pie chart
- B:** histogram
- C:** IQR
- D:** stemplot
- E:** boxplot

Feedback

A : 10

✓ Good job! The data that resulted from the survey question consisted of **one** variable (only one response was recorded for each student), and that variable was **categorical** (each student's response was one of the non-numerical categories available). The only display choice listed that is appropriate for **one categorical** variable is a pie chart.

B : 0

✗ This is not quite right. Notice that the data that resulted from the survey question were **categorical**, not quantitative (each student's response was one of the non-numerical categories available). A histogram would be the appropriate display for a quantitative variable (a list of numbers), not for a categorical variable (a list of selections from various categories). (A) is the correct answer.

C : 0

✗ This is not quite right. Remember that the IQR is one of the measures of spread for a dataset. So IQR is not a type of **graphical display**, it's a **number** (the distance of the central half of the data values). (A) is the correct answer.

D : 0

X This is not quite right. Notice that the data that resulted from the survey question were **categorical**, not quantitative (each student's response was one of the non-numerical categories available). A stemplot is a type of display for a quantitative variable (a list of numbers), not for a categorical variable (a list of selections from various categories). (A) is the correct answer.

E : 0

X This is not quite right. Notice that the data that resulted from the survey question were **categorical**, not quantitative (each student's response was one of the non-numerical categories available). A boxplot is a type of display for a quantitative variable (a list of numbers), not for a categorical variable (a list of selections from various categories). (A) is the correct answer.

Question (6)

A student survey was conducted in a major university, where data were collected from a random sample of 750 undergraduate students. One variable that was recorded for each student was the student's answer to the question: "What region of the country did you live in just prior to enrolling in this university?"

Northeast/Southeast/Northwest/Southwest/Midwest/Outside the U.S."

These data would be best displayed using which of the following?

- A:** histogram
- B:** IQR
- C:** pie chart
- D:** stemplot
- E:** boxplot

Feedback

A : 0

- X** This is not quite right. Notice that the data that resulted from the survey question were **categorical**, not quantitative (each student's response was one of the non-numerical categories available). A histogram would be the appropriate display for a quantitative variable (a list of numbers), not for a categorical variable (a list of selections from various categories). (C) is the correct answer.

B : 0

- X** This is not quite right. Remember that the IQR is one of the measures of spread for a dataset. So IQR is not a type of **graphical display**, it's a **number** (the distance of the central half of the data values). (C) is the correct answer.

C : 10

- ✓** Good job! The data that resulted from the survey question consisted of **one** variable (only one response was recorded for each student), and that variable was **categorical** (each student's response was one of the non-numerical categories available). The only display choice listed that is appropriate for **one categorical** variable is a pie chart.

D : 0

- X** This is not quite right. Notice that the data that resulted from the survey question were **categorical**, not quantitative (each student's response was one of the non-numerical categories available). A stemplot is a type of display for a quantitative variable (a list of numbers), not for a categorical variable (a list of selections from various categories). (C) is the correct answer.

E : 0

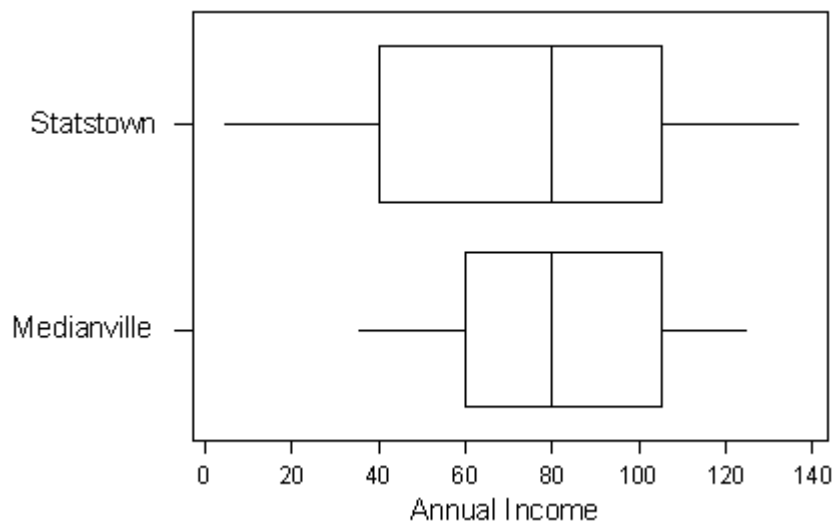
- X** This is not quite right. Notice that the data that resulted from the survey question were **categorical**, not quantitative (each student's response was one of the non-numerical categories available). A boxplot is a type of display for a quantitative variable (a list of numbers), not for a categorical variable (a list of selections from various categories). (C) is the correct answer.

Pool5

The next 4 questions relate to the same boxplot graph, shown below.

Question (7)

The boxplots below display annual incomes (in thousands of dollars) of households in two cities.



Which city has more households?

- A:** Statstown
- B:** Medianville
- C:** Both cities have the same number of households.
- D:** It is impossible to tell from the boxplots.

Feedback

A : 0

- X** This is not quite right. Remember that a boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the *income* levels for the households in the city, but not *how many households* there might be in the city. (D) is the correct answer.

B : 0

✗ This is not quite right. Remember that a boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the **income** levels for the households in the city, but not **how many households** there might be in the city. (D) is the correct answer.

C : 0

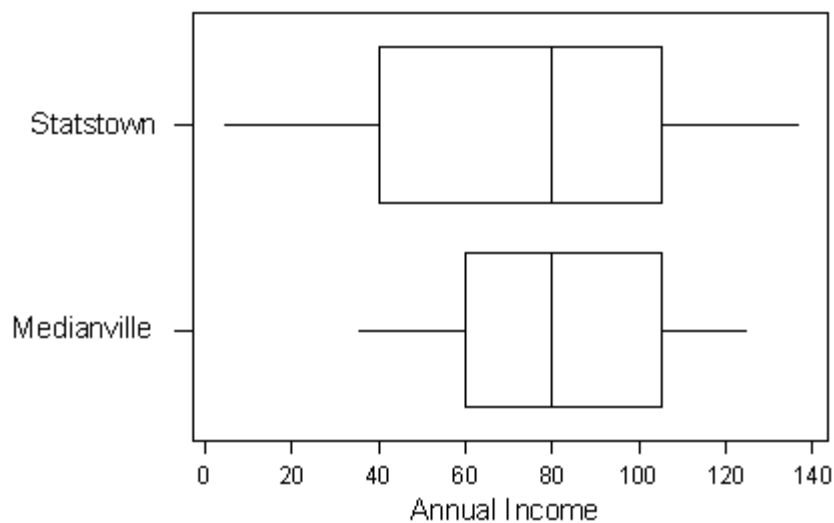
✗ This is not quite right. Remember that a boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the **income** levels for the households in the city, but not **how many households** there might be in the city. (D) is the correct answer.

D : 10

✓ Good job! A boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the **income** levels for the households in the city, but not **how many households** there might be in the city.

Question (8)

Here again are the boxplots showing annual incomes (in thousands of dollars) of households in two cities.



Which city has greater variability in income?

- A:** Statstown
- B:** Medianville
- C:** Both cities have the same variability in income.
- D:** It is impossible to tell from the boxplots.

Feedback

A : 10

✓ Good job! The city whose boxplot appears "longer" is the city whose incomes had more variability. In this case, Statstown has more variability, since we see the Statstown boxplot extends more on either side than the Medianville plot.

B : 0

✗ This is not quite right. The "length" of the boxplot is an indication of the variability of that dataset. So the city whose boxplot appears "longer" is the city whose incomes had more variability. In this case, Statstown has more variability, since we see that the 'Statstown' boxplot extends more on either side than the Medianville plot. (A) is the correct answer.

C : 0

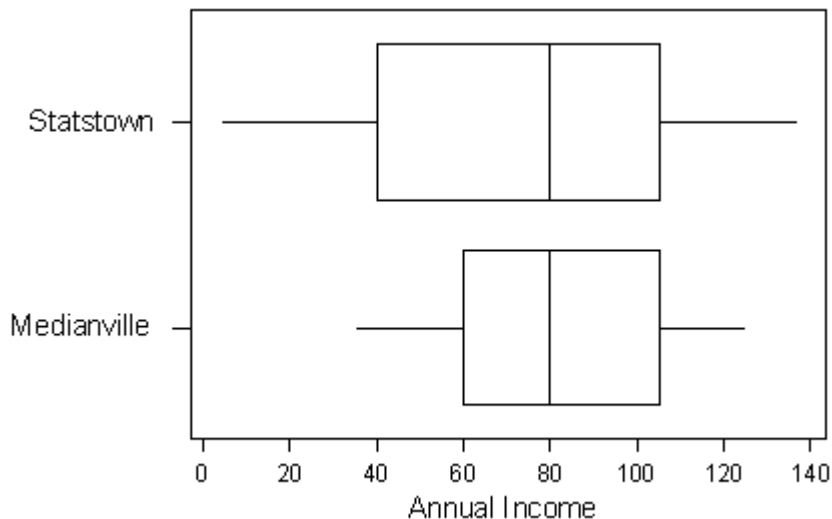
✗ This is not quite right. The "length" of the boxplot is an indication of the variability of that dataset. In this case, the two boxplots don't appear to have the same length. We see that the Statstown boxplot extends more on either side than the Medianville plot. So Statstown has more variability. (A) is the correct answer.

D : 0

X This is not quite right. The "length" of the boxplot is an indication of the variability of that dataset. So we can compare the variability of income between the two cities by comparing the side-to-side extent of the two boxplots. The city whose boxplot shows more overall length is the city with more income variation. (A) is the correct answer.

Question (9)

Here again are the boxplots showing annual incomes (in thousands of dollars) of households in two cities.



Which city has a greater percentage of households with annual incomes above \$80,000?

- A:** Statstown
- B:** Medianville
- C:** Both cities have the same percentage of households with annual incomes above \$80,000.
- D:** It is impossible to tell from the boxplots.

Feedback

A : 0

X This is not quite right. Notice the line inside each box. Remember that the line marks the **median** of the data. So in this case, **exactly 50% of the households** in each city have annual incomes of more than \$80,000. (C) is the correct answer.

B : 0

X This is not quite right. Notice the line inside each box. Remember that the line marks the **median** of the data. So in this case, **exactly 50% of the households** in each city have annual incomes of more than \$80,000. (C) is the correct answer.

C : 10

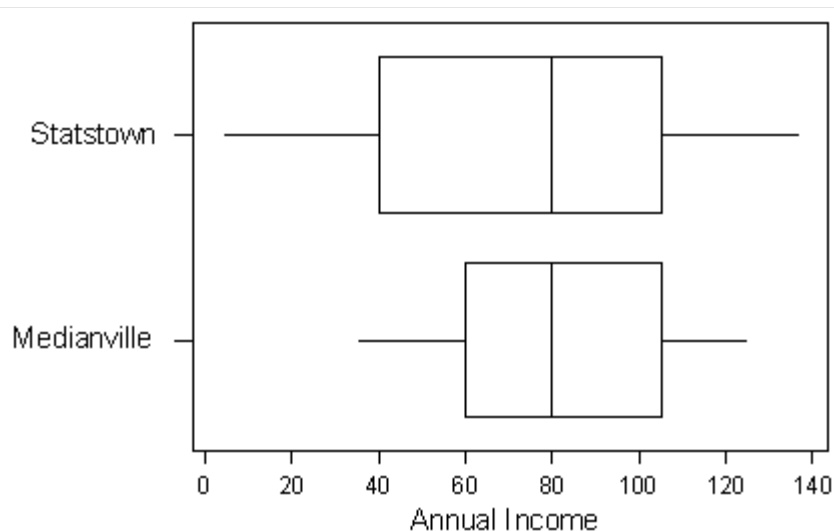
✓ Good job! In this case, the median for each city is \$80,000, because the line inside each box marks the median of the data. So in this case, **exactly 50% of the households** in each city have annual incomes of more than \$80,000.

D : 0

X This is not quite right. Notice the line inside each box. Remember that the line marks the **median** of the data, and that exactly 50% of the data points have value greater than the median. (C) is the correct answer.

Question (10)

Here again are the boxplots showing annual incomes (in thousands of dollars) for households in two cities.



Which city has a greater percentage of households with annual incomes between \$50,000 and \$80,000?

- A:** Statstown
- B:** Medianville
- C:** Both cities have the same percentage of households with annual incomes between \$50,000 and \$80,000.
- D:** It is impossible to tell from the boxplots.

Feedback

A : 0

X This is not quite right. Remember that the line in the box indicates the median. And remember that the left edge of the box indicates the first quartile. Remember that exactly a quarter of the data will be between the first quartile and the median. Now compare the display for the two cities, and rethink. (B) is the correct answer.

B : 10

✓ Good job! In fact, income between \$50,000 and \$80,000 includes less than a fourth of the households in Statstown, but **more** than a fourth of the households in Medianville, because of how the medians and the first quartiles compare for the two cities.

C : 0

X This is not quite right. Remember that the line in the box indicates the median. And remember that the left edge of the box indicates the first quartile. Remember that exactly a quarter

of the data will be between the first quartile and the median.
 ✗ Now compare the display for the two cities, and rethink. (B) is the correct answer.

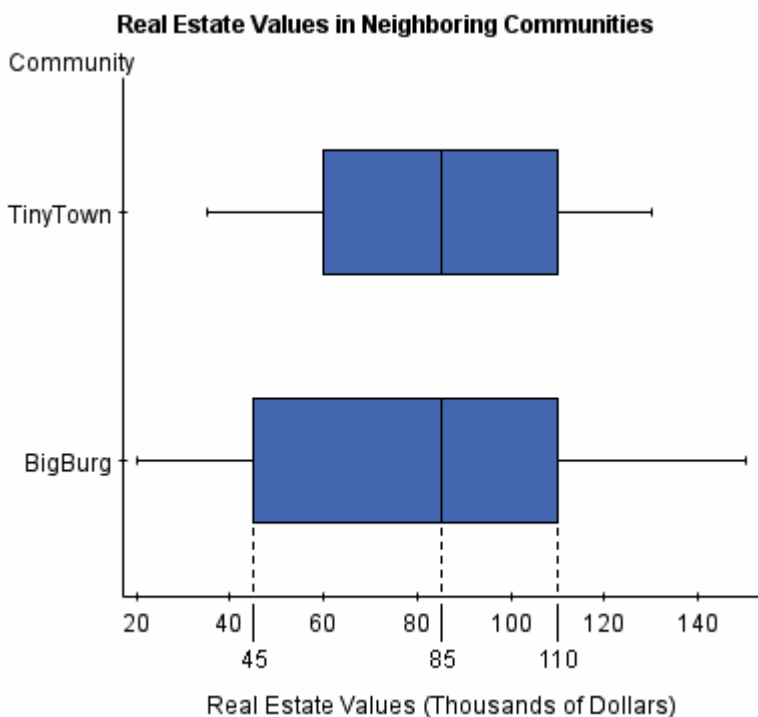
D : 0

✗ This is not quite right. Remember that the line in the box indicates the median. And remember that the left edge of the box indicates the first quartile. Remember that exactly a quarter of the data will be between the first quartile and the median. Now compare the display for the two cities, and rethink. (B) is the correct answer.

The next 4 questions relate to the same boxplot graph, shown below.

Question (11)

The boxplots below show the real estate values of single family homes in 2 neighboring cities (in thousands of dollars).




Which city has more households?


- A:** Tinytown
B: BigBurg
C: Both cities have the same number of households.
D: It is impossible to tell from the boxplots.
-

Feedback


A : 0

-  This is not quite right. Remember that a boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the **real estate values** for the homes in the city, but not **how many homes** there might be in the city. (D) is the correct answer.


B : 0

-  This is not quite right. Remember that a boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the **real estate values** for the homes in the city, but not **how many homes** there might be in the city. (D) is the correct answer.

C : 0

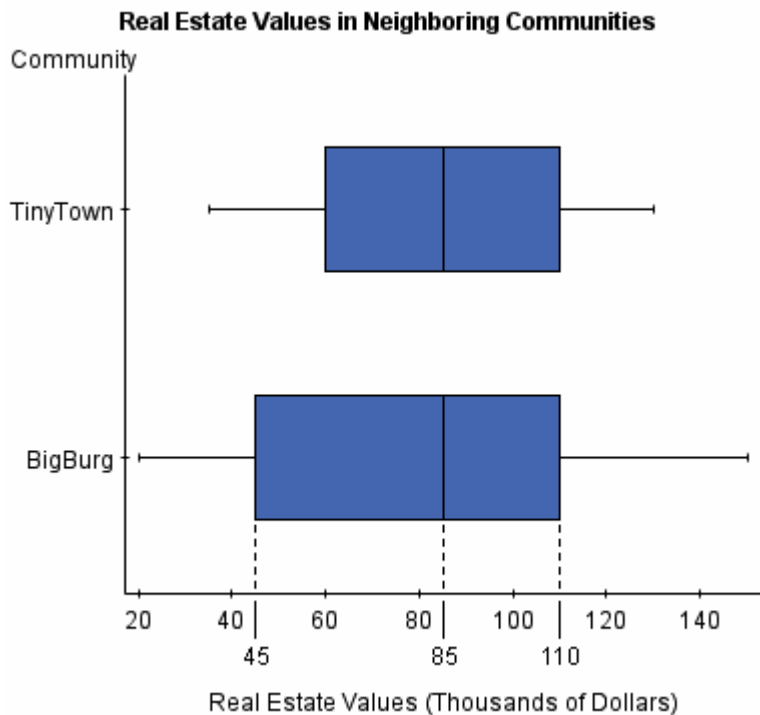
-  This is not quite right. Remember that a boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the **real estate values** for the homes in the city, but not **how many homes** there might be in the city. (D) is the correct answer.

D : 10

-  Good job! A boxplot only displays the five-number summary of a dataset. A boxplot doesn't indicate how many data values might have been in the dataset. So, in this case, the boxplots only indicate the **real estate values** for the homes in the city, but not **how many homes** there might be in the city. (D) is the correct answer.

Question (12)

Here again are the boxplots showing the real estate values of single family homes in 2 neighboring cities (in thousands of dollars).



Which city has greater variability in real estate values?

- A:** TINYtown
- B:** Bigburg
- C:** Both cities have the same variability in real estate values.
- D:** It is impossible to tell from the boxplots.

Feedback

A : 0

X This is not quite right. The "length" of the boxplot is an indication of the variability of that dataset. So the city whose boxplot appears "longer" is the city whose real estate values had more variability. In this case, Bigburg has more variability, since we see that the Bigburg boxplot extends more on either side than the TINYtown plot. (B) is the correct answer.

B : 10

✓ Good job! The city whose boxplot appears "longer" is the city whose real estate values had more variability. In this case, Bigburg has more variability, since we see that the Bigburg boxplot extends more on either side than the Tinytown plot.

C : 0

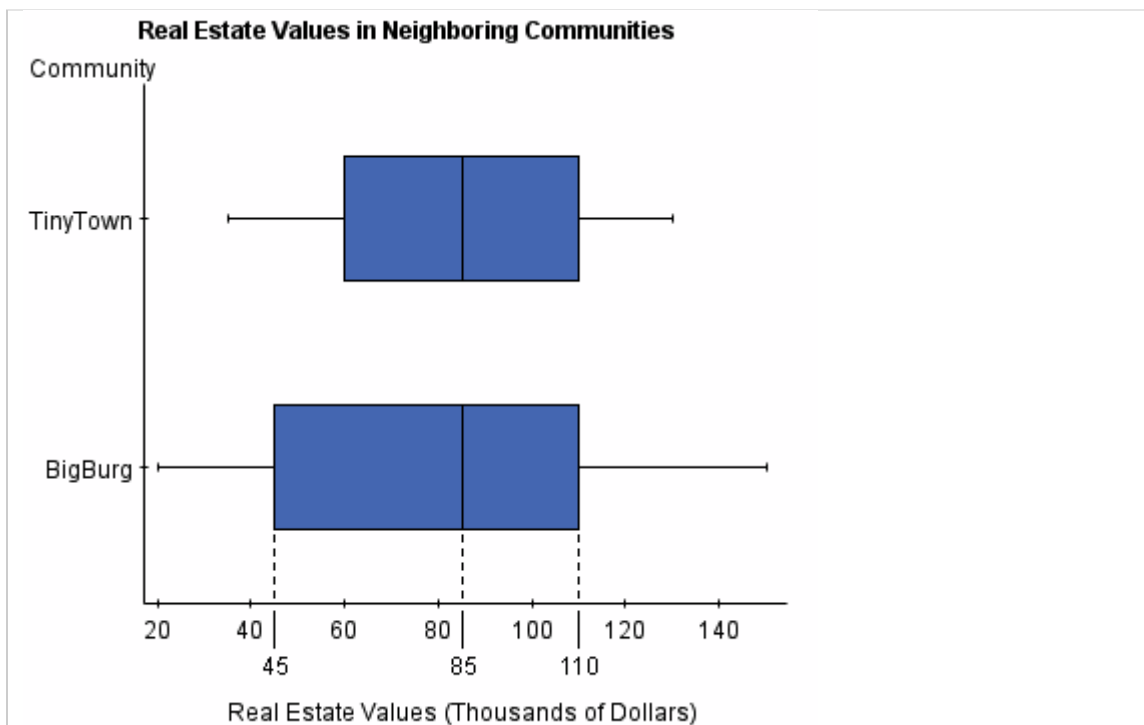
✗ This is not quite right. The "length" of the boxplot is an indication of the variability of that dataset. So the city whose boxplot appears "longer" is the city whose real estate values had more variability. In this case, Bigburg has more variability, since we see that the Bigburg boxplot extends more on either side than the Tinytown plot. (B) is the correct answer.

D : 0

✗ This is not quite right. The "length" of the boxplot is an indication of the variability of that dataset. So the city whose boxplot appears "longer" is the city whose real estate values had more variability. In this case, Bigburg has more variability, since we see that the Bigburg boxplot extends more on either side than the Tinytown plot. (B) is the correct answer.

Question (13)

Here again are the boxplots showing the real estate values of single family homes in 2 neighboring cities (in thousands of dollars).



Which city has the greater percentage of households with real estate values above \$85,000?

- A:** Tinytown
- B:** Bigburg
- C:** Both cities have the same percentage of households with real estate values above \$85,000.
- D:** It is impossible to tell from the boxplots.

Feedback

A : 0

X This is not quite right. Notice the line inside each box. Remember that the line marks the **median** of the data. So in this case, **exactly 50% of the homes** in each city have real estate values of more than \$85,000. (C) is the correct answer.

B : 0

X This is not quite right. Notice the line inside each box. Remember that the line marks the **median** of the data. So in this case, **exactly 50% of the homes** in each city have real estate values of more than \$85,000. (C) is the correct answer.

C : 10

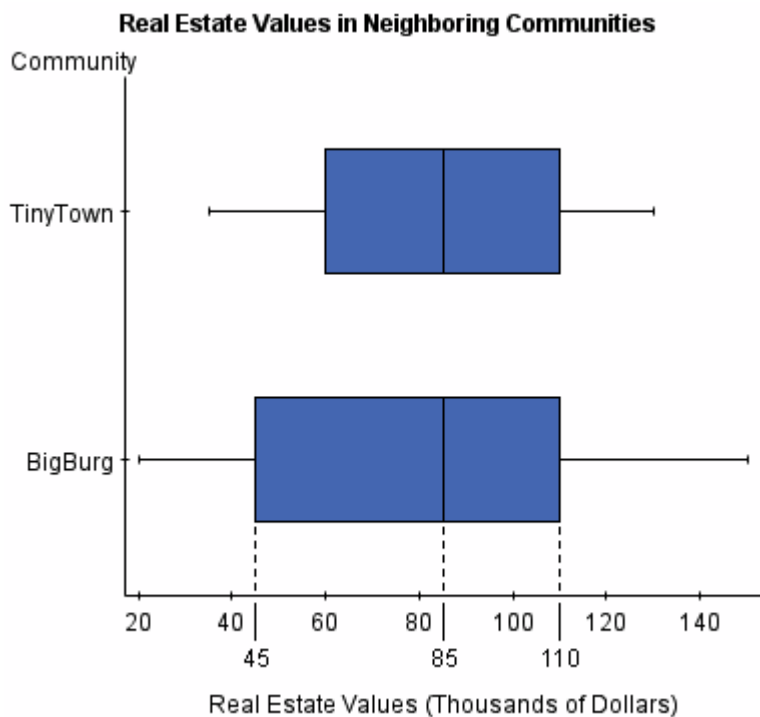
✓ Good job! In this case, the median for each city is \$85,000, because the line inside each box marks the median of the data. So in this case, **exactly 50% of the homes** in each city have real estate values of more than \$85,000.

D : 0

✗ This is not quite right. Notice the line inside each box. Remember that the line marks the **median** of the data, and that exactly 50% of the data points have value greater than the median. (C) is the correct answer.

Question (14)

Here again are the boxplots showing the real estate values of single family homes in 2 neighboring cities (in thousands of dollars).



Which city has a greater percentage of homes with real estate values between \$55,000 and \$85,000?

A: Tintytown

B: BigBurg

C: Both cities have the same percentage of homes with real estate values between \$55,000 and \$85,000.

D: It is impossible to tell from the boxplots.

Feedback

A : 10

✓ Good job! In fact, values between \$55,000 and \$85,000 include less than a fourth of the homes in Bigburg, but **more** than a fourth of the households in Tintytown, because of how the medians and the first quartiles compare for the two cities.

B : 0

✗ This is not quite right. Remember that the line in the box indicates the median. And remember that the left edge of the box indicates the first quartile. Remember that exactly a quarter of the data will be between the first quartile and the median. Now compare the display for the two cities, and rethink. (A) is the correct answer.

C : 0

✗ This is not quite right. Remember that the line in the box indicates the median. And remember that the left edge of the box indicates the first quartile. Remember that exactly a quarter of the data will be between the first quartile and the median. Now compare the display for the two cities, and re-think. (A) is the correct answer.

D : 0


✗ This is not quite right. Remember that the line in the box indicates the median. And remember that the left edge of the box indicates the first quartile. Remember that exactly a quarter of the data will be between the first quartile and the median. Now compare the display for the two cities, and rethink. (A) is the correct answer.

Please answer the question below. Your response will not be graded, but will be available for your instructor to read.

What determines which numerical measures of center and spread are appropriate for describing a given distribution of a quantitative variable? Which measures will you use in each case?

Feedback

* : 0

 Thank you for your response. It has been recorded and will be available for your instructor to read.

 You did not enter a response for this question.